CHAPTER XIII Information Theory and Art

SOME MONTHS AGO when a competent modern composer and professor of music visited the Bell Laboratories, he was full of the news that musical sounds and, in fact, whole musical compositions can be reduced to a series of numbers. This was of course old stuff to us. By using pulse code modulation, one can represent any electric or acoustic wave form by means of a sequence of sample amplitudes.

We had considered something that the composer didn't appreciate. In order to represent fairly high-quality music, with a band width of 15,000 cycles per second, one must use 30,000 samples per second, and each one of these must be specified to an accuracy of perhaps one part in a thousand. We can do this by using three decimal digits (or about ten binary digits) to designate the amplitude of each sample.

A composer could exercise complete freedom of choice among sounds simply by specifying a sequence of 30,000 three-digit decimal numbers a second. This would allow him to choose from among a number of twenty-minute compositions which can be written as 1 followed by 108 million 0's—an inconceivably large number. Putting it another way, the choice he could exercise in composing would be 300,000 bits per second.

Here we sense what is wrong. We have noted that by the fastest demonstrated means, that is, by reading lists of words as rapidly as possible, a human being demonstrates an information rate of

no more than 40 bits per second. This is scarcely more than a ten-thousandth of the rate we have allowed our composer.

Further, it may be that a human being can make use of, can appreciate, information only at some rate even less than 40 bits per second. When we listen to an actor, we hear highly redundant English uttered at a rather moderate speed.

The flexibility and freedom that a composer has in expressing a composition as a sequence of sample amplitudes is largely wasted. They allow him to produce a host of "compositions" which to any human auditor will sound indistinguishable and uninteresting. Mathematically, white Gaussian noise, which contains all frequencies equally, is the epitome of the various and unexpected. It is the least predictable, the most original of sounds. To a human being, however, all white Gaussian noise sounds alike. Its subtleties are hidden from him, and he says that it is dull and monotonous.

If a human being finds monotonous that which is mathematically most various and unpredictable, what does he find fresh and interesting? To be able to call a thing new, he must be able to distinguish it from that which is old. To be distinguishable, sounds must be to a degree familiar.

We can tell our friends apart, we can appreciate their particular individual qualities, but we find much less that is distinctive in strangers. We can, of course, tell a Chinese from our Caucasian friends, but this does not enable us to enjoy variety among Chinese. To do that we have to learn to know and distinguish among many Chinese. In the same way, we can distinguish Gaussian noise from Romantic music, but this gives us little scope for variety, because all Gaussian noise sounds alike to us.

Indeed, to many who love and distinguish among Romantic composers, most eighteenth-century music sounds pretty much alike. And to them Grieg's *Holberg Suite* may sound like eighteenth-century music, which it resembles only superficially. Even to those familiar with eighteenth-century music, the choral music of the sixteenth century may seem monotonous and undistinguishable. I know, too, that this works in reverse order, for some partisans of Mozart find Verdi monotonous, and to those for whom Verdi affords tremendous variety much modern music sounds like undistinguishable noise.

Of course a composer wants to be free and original, but he also

wants to be known and appreciated. If his audience can't tell one of his compositions from another, they certainly won't buy recordings of many different compositions. If they can't tell his compositions from those of a whole school of composers, they may be satisfied to let one recording stand for the lot.

How, then, can a composer make his compositions distinctive to an audience? Only by keeping their entropy, their information rate, their variety within the bounds of human ability to make distinctions. Only when he doles his variety out at a rate of a very few bits per second can he expect an audience to recognize and appreciate it.

Does this mean that the calculating composer, the information-theoretic composer so to speak, will produce a simple and slow succession of randomly chosen notes? Of course not, not any more than a writer produces a random sequence of letters. Rather, the composer will make up his composition of larger units which are already familiar in some degree to listeners through the training they have received in listening to other compositions. These units will be ordered so that, to a degree, a listener expects what comes next and isn't continually thrown off the track. Perhaps the composer will surprise the listener a bit from time to time, but he won't try to do this continually. To a degree, too, the composer will introduce entirely new material sparingly. He will familiarize the listener with this new material and then repeat the material in somewhat altered forms.

To use the analogy of language, the composer will write in a language which the listener knows. He will produce a well-ordered sequence of musical words in a musically grammatical order. The words may be recognizable chords, scales, themes, or ornaments. They will succeed one another in the equivalents of sentences or stanzas, usually with a good deal of repetition. They will be uttered by he familiar voices of the orchestra. If he is a good composer, he will in some way convey a distinct and personal impression to the skilled listener. If he is at least a skillful composer, his composition will be intelligible and agreeable.

Of course, none of this is new. Those quite unfamiliar with information theory could have said it, and they have said it in other words. It does seem to me, however, that these facts are particularly pertinent to a day in which composers, and other artists as well, are faced with a multitude of technical resources which are tempting, exasperating, and a little frightening.

Their first temptation is certainly to choose too freely and too widely. M. V. Mathews of the Bell Laboratories was intrigued by the fact that an electronic computer can create any desired wave form in response to a sequence of commands punched into cards. He devised a program such that he could specify one note by each card as to wave form, duration, pitch, and loudness. Delighted with the freedom this afforded him, he had the computer reproduce rapid rhythmic passages of almost unplayable combinations, such as three notes against four with unusual patterns of accent. These ingenious exercises sounded, simply, chaotic.

Very skillful composers, such as Varèse, can evoke an impression of form and sense by patching together all sorts of recorded and modified sounds after the fashion of *musique concrète.* Many appealing compositions utilizing electronically generated sounds have already been produced. Still, the composer is faced with difficulties when he abandons traditional resources.

The composer can choose to make his compositions much simpler than he would if he were writing more conventionally, in order not to lose his audience. Or he and others can try to educate an audience to remember and distinguish among the new resources of which they avail themselves. Or the composer can choose to remain unintelligible and await vindication from posterity. Perhaps there are other alternatives; certainly there are if the composer has real genius.

Does information theory have anything concrete to offer concerning the arts? I think that it has very little of serious value to offer except a point of view, but I believe that the point of view may be worth exploring in the brief remainder of this chapter.

In Chapters III, VI, and XII we considered language. Language consists of an alphabet or vocabulary of words and of grammatical rules or constraints concerning the use of words in grammatical text. We learned to distinguish between the features of text which are dictated by the vocabulary and the rules of grammar and the actual choice exercised by the writer or speaker. It is only this element of choice which contributes to the average amount of

information per word. We saw that Shannon has estimated this to be between 3.3 and 7.2 bits per word. It must also be this choice which enables a writer or speaker to convey meaning, whatever that may be.

The vocabulary of a language is large, although we have seen in Chapter XII that a comparatively few words make up the bulk of any text. The rules of grammar are so complicated that they have not been completely formulated. Nonetheless, most people have a large vocabulary, and they know the rules of grammar in the sense that they can recognize and write grammatical English.

It is reasonable to assume a similarly surprisingly large knowledge of musical elements and of relations among them on the part of the person who listens to music frequently, attentively, and appreciatively. Of course, it is not necessary that the listener be able to formulate his knowledge for him to have it, any more than the writer of grammatical English need be able to formulate the rules of English grammar. He need not even be able to write music according to the rules, any more than a mute who understands speech can speak. He can still in some sense *know* the rules and make use of his knowledge in listening to music.

Such a knowledge of the elements and rules of the music of a particular nation, era, or school is what I have referred to as "knowing the language of music" or of a style of music. However much the rules of music may or may not be based on physical laws, a knowledge of a language of music must be acquired by years of practice, just as the knowledge of a spoken language is. It is only by means of such a knowledge that we can distinguish the style and individuality of a composition, whether literary or musical. To the untutored ear, the sounds of music will seem to be examples chosen not from a restricted class of learned sounds but from all the infinity of possible sounds. To the untutored ear, the mechanical workings of the rules of music will seem to represent choice and variety. Thus, the apparent complexity of music will overwhelm the untutored auditor or the auditor familiar only with some other language of music.

We should note that we can write sense while violating the rules of grammar to a degree (me heap big injun). We might liken the intelligibility of this sentence to an English-speaking person to our

ability to appreciate music which is somewhat strange but not entirely foreign to our experience. We should also note that we can write nonsense while obeying the rules of grammar carefully (the alabaster word spoke silently to the purple). It is to this second possibility to which I wish to address myself in a moment. I will remark first, however, that while one can of course both write sense and obey the rules while doing so, he often exposes his inadequacies to the public gaze by thus being intelligible.

It is no news that we can dispense with sense almost entirely while retaining a conventional vocabulary and some or many rules. Thus, Mozart provided posterity with a collection of assorted, numbered bars in 3/8 time, together with a set of rules (Koechel 294D). By throwing dice to obtain a sequence of random numbers and choosing successive bars by means of the rules, even the nonmusical amateur can "compose" an almost endless number of little waltzes which sound like somewhat disorganized Mozart. An example is shown in Figure XIII-1. Joseph Haydn, Maximilian Stadler, and Karl Philipp Emanuel Bach are said to have produced similar random music. In more recent times, John Cage has used random processes in the choice of sequences of notes.

In ignorance of these illustrious predecessors, in 1949 M. E. Shannon (Claude Shannon's wife) and I undertook the composition of some very primitive statistical or stochastic music. First we made a catalog of allowed chords on roots I–VI in the key of C. Actually, the catalog covered only root I chords; the others were

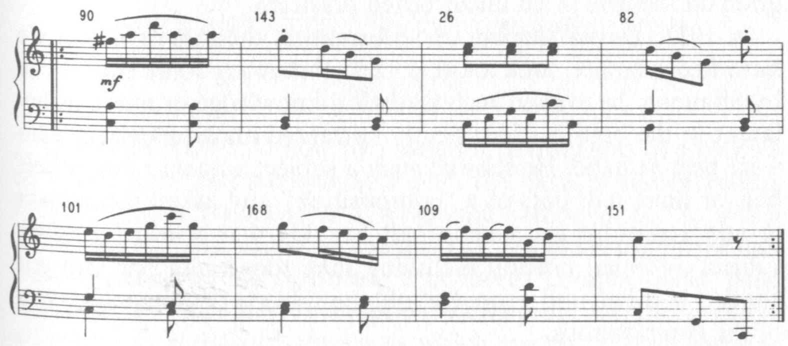

Fig. XIII-1

derived from these by rules. By throwing three specially made dice and by using a table of random numbers, a number of compositions were produced.

In these compositions, the only rule of chord connection was that two succeeding chords have a common tone in the same voice. This let the other voices jump around in a wild and rather unsatisfactory manner. It would correspond to the use of a simple and consistent but incorrect digram probability in the construction of synthetic text, as illustrated in Chapter III.

While the short-range structure of these compositions was very primitive, an effort was made to give them a plausible and reasonably memorable, longer-range structure. Thus, each composition consisted of eight measures of four quarter notes each. The long-range structure was attained by making measures 5 and 6 repeat measures 1 and 2, while measures 3 and 4 differed from measures 7 and 8. Thus, the compositions were primitive rondos. Further, it was specified that chords 1, 16, and 32 have root I and chords 15 and 31 have either root IV or root V, in order to give the effect of a cadence.

Although the compositions are formally rondos, they resemble hymns. I have reproduced one as Figure XIII-2. As all hymns should have titles and words, I have provided these by nonrandom means. The other compositions sound much like the one given. Clearly, they are all by the same composer. Still, after a few hearings they can be recognized as different. I have even managed to grow fond of them through hearing them too often. They must grate on the ears of an uncorrupted musician.

In 1951, David Slepian, an information theorist of whom we have heard before, took another tack. Following some early work by Shannon, he evoked such statistical knowledge of music as lay latent in the breasts of musically untrained mathematicians who were near at hand. He showed such a subject a quarter bar, a half bar, or three half bars of a "composition" and asked the subject to add a sensible succeeding half bar. He then showed another subject an equal portion including that added half bar and got another half bar, and so on. He told the subjects the intended styles of the compositions.

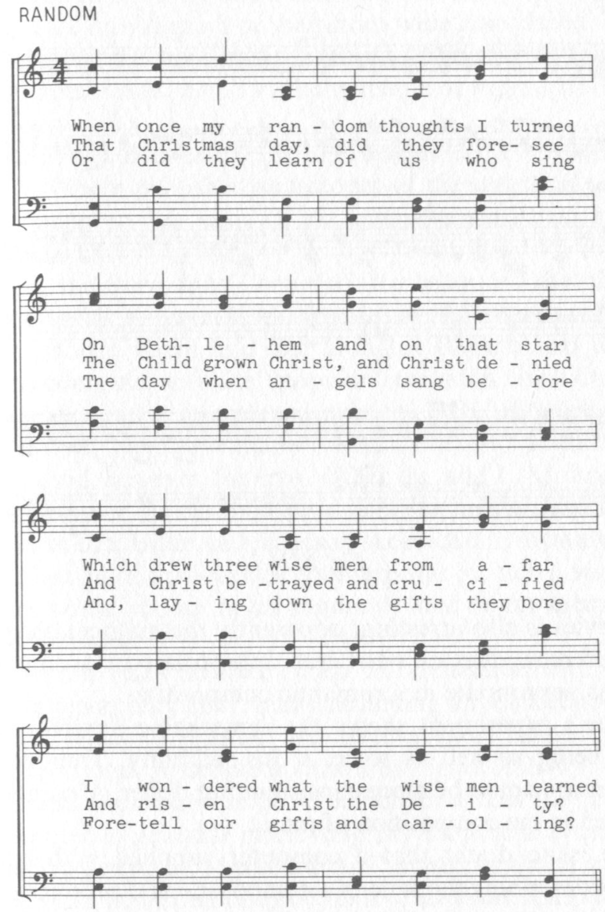

Fig. XIII-2

In Figure XIII-3, I show two samples: a fragment of a chorale in which each half bar was added on the basis of the preceding half bar and a fragment of a "romantic composition," in which each half bar was added on the basis of the preceding three half bars. It seems to me surprising that these "compositions" hang together as well as they do, despite the inappropriate and inadmissible chords and chord sequences which appear. The distinctness

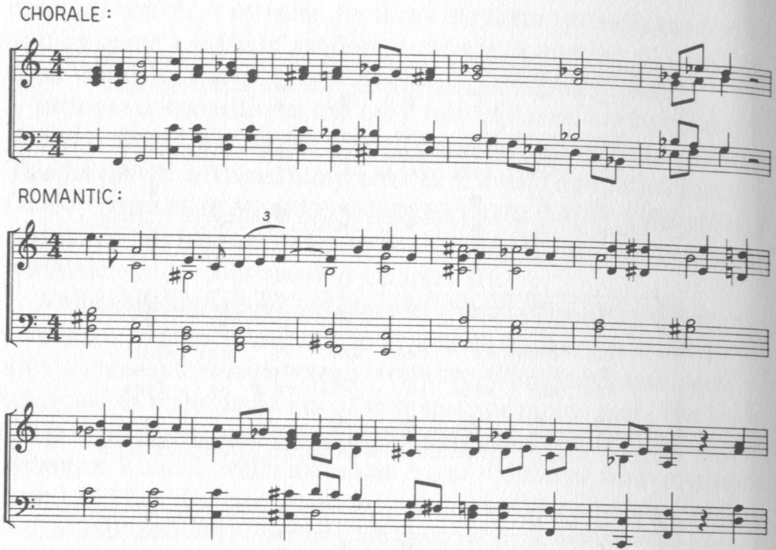

Fig. XIII-3

of the styles is also arresting; apparently the mathematicians had quite different ideas of what was appropriate in a chorale and what was appropriate in a romantic composition.

Slepian's experiment shows the remarkable flexibility of the human being as well as some of his fallibility. True stochastic processes are apt to be more consistent but duller. A number have been used in the composition of music.

There is no doubt that a computer supplied with adequate statistics describing the style of a composer could produce random music with a recognizable similarity to a composer's style. The nursery-tune style demonstrated by Pinkerton and the diversity of styles evoked by Hiller and Isaacson, which I will describe presently, illustrate this possibility.

In 1956, Richard C. Pinkerton published in the *Scientific American* some simple schemes for writing tunes. He showed how a note could be chosen on the basis of its probability of following the particular preceding note and how the probabilities changed with respect to the position of the note in the bar. Using probabilities derived from nursery tunes, he computed the entropy per note,

which he found to be 2.8 bits. I feel sure that this is quite a bit too high, because only digram probabilities were considered. He also presented a simple finite-state machine which could be used to generate banal tunes, much as the machine of Figure III-1 generates "sentences."

In 1957, F. B. Brooks, Jr., A. L. Hopkins, Jr., P. G. Neumann, and W. V. Wright published an account of the statistical composition of music on the basis of an extensive statistical study of hymn tunes.

In 1956, the Burroughs Corporation announced that they had used a computer to generate music, and, in 1957, it was announced that Dr. Martin Klein and Dr. Douglas Bolitho had used the Datatron computer to write "popular" melodies. Jack Owens set words to one, and it was played over the ABC network as *Push Button Bertha*. No doubt many others have done similar things.

It remained, however, for L. A. Hiller, Jr., and L. M. Isaacson of the University of Illinois to make a really serious experiment with computer music. Hiller and Isaacson succeeded in formulating the rules of four-part, first-species counterpoint in such a way that a computer could choose notes randomly and reject them if they violated the rules.

Because the rules involve, except in connection with the concluding cadence, only direct relations among three successive notes, the music tends to wander, but over a short range it sounds surprisingly good. A sample is shown in Figure XIII-4.[1]

Hiller and Isaacson went on to demonstrate that they could use the computer to generate interesting rhythmic and dynamic patterns and to generate "Markoff-chain" music, in which successive note selection depended on probability functions computed from tables derived from various considerations of overtones or harmonics. In this case they generated a coda according to a simple prescription.

As it stands, this music, which was brought together and published as the *Illiac Suite for String Quartet,* has a good deal of local structure but is weak and wandering as a whole. The imposition

[1] Reproduced from L. A. Hiller, Jr., and L. M. Isaacson, *Illiac Suite for String Quartet,* New Music, 1957, by permission of Theodore Presser Company, Bryn Mawr, Pa.

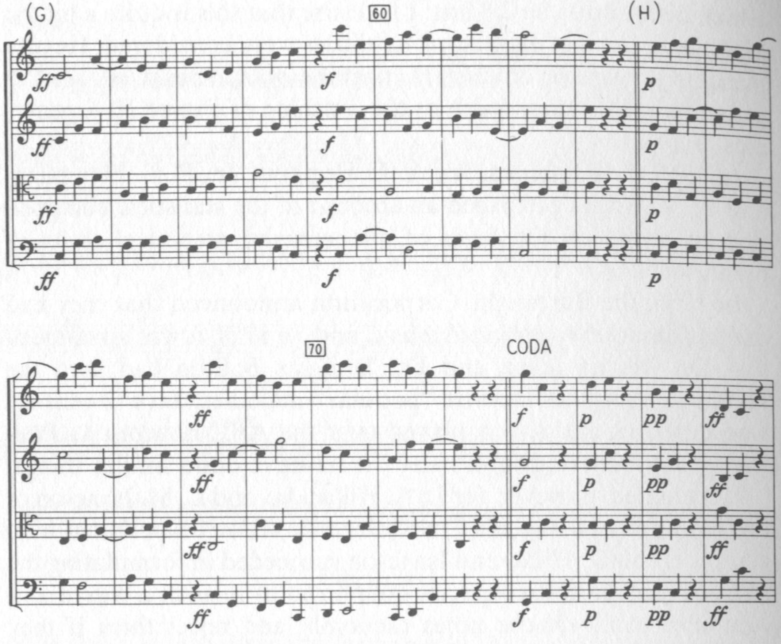

Fig. XIII-4

of some simple pattern or repetition might have helped considerably. This could be of a strictly deterministic nature, as in the case of the prescribed repetitions in a rondo, or it could be of the nature of Chomsky's grammar, which we have considered in Chapter VI. It is clear, however, that it is foolish to try to attain long-range structure simply by relating a note to the immediately preceding notes by digram, trigram, and higher probabilities. The relation must be among *parts* of the composition, not simply among notes.

The work of Hiller and Isaacson does demonstrate conclusively that a computer can take over many musical chores which only human beings had been able to do before. A composer, and especially an unskilled composer, might very well rely on a computer for much routine musical drudgery. The composer could merely guide the main pattern of the composition and let the computer fill in details of harmony and counterpoint, according to a specification of style or period. Further, the computer could

be used to try out proposed new rules of composition, such as new rules of counterpoint or harmony, with whose use and consequences the composer might have little experience and familiarity.

In these days we hear that cybernetics will soon give us machines which learn. If they learn in a complicated enough sense of the word, why couldn't they learn what we like, even when we don't know ourselves? Thus, by rewarding or punishing a computer for the success or failure of its efforts, we might so condition the computer that when we pressed a button marked Spanish, classical, rock-and-roll, sweet, etc., it would produce just what we wanted in connection with the terms. Such thoughts are intriguing, but they are of course nonsense in our day and will probably remain so for a long time to come.

Music is not all of art. I began with music because it offers an apt means for illustrating in an unusual context some ideas derived from information theory. We could just as well draw our illustrations from the use of language. Indeed, experiments with the stochastic production of text have been perhaps more widely cultivated than experiments with music.

A professor at the Grand Academy of Lagoda showed Captain Lemuel Gulliver a word frame consisting of lettered blocks mounted on shafts. The professor turned these at random and sought new wisdom in the patterns of letters which appeared.

Here we see just the wrong application of a stochastic process in the generation of text. Certainly, this will not give us new knowledge. Who would take the uncorroborated word of a random process? There are all too many unsubstantiated statements available; what we need to know is what is so and what isn't.

Nonetheless, a stochastic process can produce some interesting effects. In Chapter III we noted Shannon's approximations to English text. These were made by using letter digram and trigram frequencies and a table of random numbers. We have seen that they contain some interesting "words."

To me, *deamy* has a pleasant sound; I would take "it's a deamy idea" in a complimentary sense. On the other hand, I'd hate to be denounced as *ilonasive.* I would not like to be called *grocid;* perhaps it reminds me of gross, groceries, and gravid. *Pondenome,* whatever it may be, is at least dignified.

I repeat Shannon's second-order word approximation here:

The head and in frontal attack on an English writer that the character of this point is therefore another method for the letters that the time who ever told the problem for an unexpected.

I find this disquieting. I feel that the English writer is in mortal peril, yet I cannot come to his aid because the latter part of the message is garbled.

In seeking less garbled material, as I noted in Chapter VI, I wrote three grammatically connected words in a column from the top down on a slip of paper. I showed them to a friend, asked him to make up a sentence in which they occurred, and then to add the next word in this sentence. I then folded over the top word of the four I now had and showed the visible three to another friend and got another word from him. After canvassing twenty friends, I had the following:

When morning broke after an orgy of wild abandon he said her head shook vertically aligned in a sequence of words signifying what . . .

Later examples are:

One day when I went to what was Dionysus thinking of women without men go off half way decent impression . . .

I forget whether he went on and on. Finally he stipulated that this must stop immediately after this. The last time I saw him when she lived. It happened one frosty look of trees waving gracefully against the wall. You never can . . .

We see that a seeming order of meaning persists over groups of far more than four words. Eventually, however, the text wanders. The long-range wandering is of course attributable to the fact that there is no long-range, persistent purpose or meaning guiding the choice of words. We sometimes see a similar quality in the utterances of schizophrenics. W. H. Hudson illustrates a more gradual wandering admirably by a character in his *The Purple Land:* Uncle Anselmo never gets to the end of a story because he continually wanders on to new ground.

One can add a certain amount of long-range order by writing, in view, at the bottom of the slip of paper to which people add

words a title which indicates what the passage is supposed to be about.

Dr. Donald A. Dunn of the Stanford Electronics Laboratory has kindly supplied me with some examples in which the person adding a word saw only one preceding word, together with the title:

Men and Women

Eve loved intensely sentimental or not sufficient tonight wherever you may die before yesterday again and whatever m'love misbehaves.

The seduced are compatible unusual family life seemed wonderful experience for tenderness forever yours.

Orphans frequently visited his promiscuity and infidelity despite hate and love for tomorrow sex ain't nothing.

In the following examples, which were produced at the Bell Laboratories, the person adding a word saw three preceding words as well as the title:

About Life

Life has many good and wise men seldom condemn halfwits lightly! You wonder why not. Human feelings but savage tribes found . . .

Engineers

It is frequently said that they knew why forces might affect salaries. However, all scientists can't imagine . . .

Housecleaning

First empty the furniture of the master bedroom and bath. Toilets are to be washed after polishing doorknobs the rest of the room. Washing windows semiannually is to be taken by small aids such as husbands are prone to omit soap powder.

Murder Story

When I killed her I stabbed Claude between his powerful jaws clamped tightly together. Screaming loudly despite fatal consequences in the struggle for life ebbing as he coughed hollowly spitting blood from his ears.

I think that it is hard to read such material without amusement. I feel a little admiration as well. I would never write, "It happened one frosty look of trees waving gracefully against the wall." I almost wish I could. Poor poets endlessly rhyme love with dove,

and they are constrained by their highly trained mediocrity *never* to produce a good line. In some sense, a stochastic process can do better; it at least has a chance. I wish I had hit on *deamy,* but I never would have.

Will a computer produce text of any literary merit by means of grammatical rules and a sequence of random numbers? It might produce fresh and amusing "words" and amusing short passages of some shock value. One can of course imagine a machine designed to write detective novels and equipped with settings for hard-boiled, puzzle, character, suspense, and so on, but such a device seems to me to be very far away.

The visual arts can be used to illustrate the same points which have been made in connection with music and language. A completely random visual pattern, like a completely random acoustic wave or a completely random sequence of letters, is mathematically the most surprising, the least predictable of all possible patterns. Alas, a completely random pattern is also the dullest of all patterns, and to a human being one random pattern looks just like another. Figure XIII-5, which is an array of 10,000 randomly black or white dots, illustrates this.

Bela Julesz, who works in the field of perception, caused an

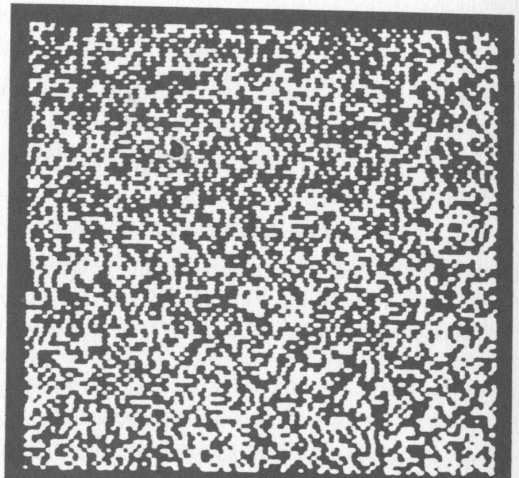

Fig. XIII-5

electronic computer to produce this random array of dots as a part of his studies of stereoscopic vision and of the meaning of pattern. He also programmed the computer to remove some of the randomness from such a random pattern. He did this by making the computer examine successively various sets of five points located at the tips and at the center of an *X,* as shown by the points marked *X* in Figure XIII-6 (other points are marked O). If the center point was the same (black or white) as either points 1 and 4 or points 2 and 3, it was changed (from black to white or from white to black). This tends to remove any black or white diagonals, except when points 1 and 4 are black and points 2 and 3 are white or vice versa.

As we can see from Figure XIII-7, making a pattern less random in this way alters and improves its appearance profoundly. An unpredictable (random) component is desirable for the sake of variety or surprise, but some orderliness is necessary if a pattern is to be pleasing.

This exploitation of both order and randomness is in fact old to art. The kaleidoscope offers a charming effect by giving to a random arrangement of bits of colored glass a sixfold symmetry.

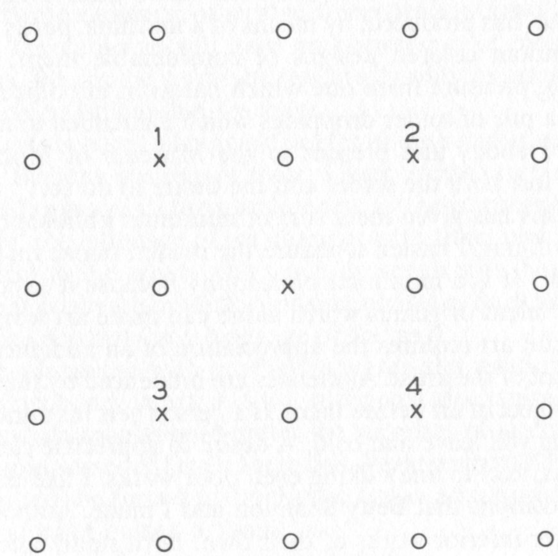

Fig. XIII-6

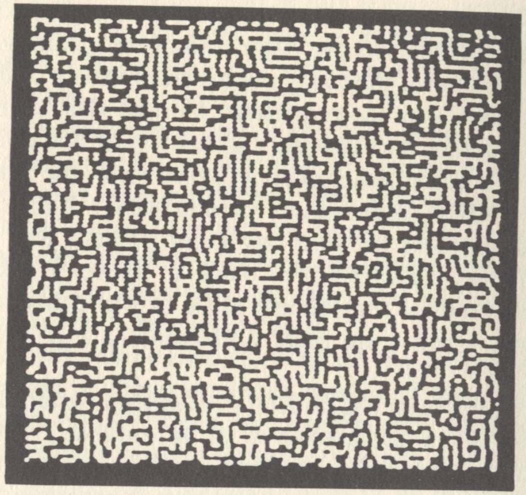

Fig. XIII-7

Many years ago Marcel Duchamp, who painted *Nude Descending a Staircase,* allowed a number of threads to fall on pieces of black cloth and then framed and preserved them. Jean Tinguley, the Swiss artist, has produced, by means of a machine, partly ordered, partly random colored designs of considerable merit; I derive continuing pleasure from one which hangs in my office. I saved for years a pile of solder droppings which I intended to mount on a block of ebony and present to the Museum of Modern Art. Finally, I lost both the solder and the desire to do so.

All of this has given me a sort of minimum philosophy of art, which I will not, I hasten to assure the reader, blame on information theory. It is a minimum philosophy because it says nothing about the talent or genius which alone can make art worth while.

Successful art requires the appreciation of an audience as well as the talent of the artist. Audiences are influenced by things other than the object of art before them. If a person sets his mind against it, anything will leave him cold. A desire to appreciate can, on the other hand, lead to one's liking even poor works. I like the hymn-like compositions that Betty Shannon and I made. Authors sometimes prefer inferior works of their own. Both small coteries and large groups can be led to appreciate sincerely things which are for

a time the fashion but which have little long-range appeal and which probably have little merit.

Among other things, audiences want to have a sense of authorship, a sense of an individual, in connection with works of art. To bring appreciation to an artist, his work must have enough consistency so that it is recognizable as his. How let down the sincere appreciator must be if he always has to look at the label or wait for the announcer in order to know that the painting or music is the product of his favorite artist.

Suppose that one artist had actually produced in succession the masterpieces we now accept as the works of a number of great artists with diverse styles, long before the artists lived. This would astonish us, but we could scarcely appreciate him as an artist, however much we might admire the individual paintings. Picasso is eminently recognizable, but he is disquieting. He has been skillful in many styles, and yet he escapes our final judgment by going from one style to another. How much easier it is to appreciate Matisse.

To be appreciated by an audience, art must be intelligible to the audience. Even a good joke in Chinese will amuse few Americans, and certainly ten jokes in Chinese will be no more amusing than one. To a degree, to be appreciated art must be in a language familiar to the audience; otherwise no matter how great the variety may be, the audience will have an impression of monotony, of sameness. We can be surprised repeatedly only by constrast with that which is familiar, not by chaos.

Some artists adopt a language taught to their audience by earlier masters. Brahms was one of these. Other artists teach something of a new language to their audiences, as the impressionists did. Certainly, the language of art changes with time, and we should be grateful to the artists who teach us new words. However, we should not doubt the originality of such artists as Bach and Handel, who spoke ringingly in a language of the past.

While a language with intelligible words and relations between words is necessary in art, it is not sufficient. Mechanical sameness is dull and disappointing. I prefer the surprises of stochastic prose to the vapid verses of Owen Meredith. Perhaps in some age of bad art, man will be forced to stochastic art as an alternative to the stale product of human artisans.

So much for information theory and art.

audience need (consumer, market repute)

the best artist is the best self editor

CHAPTER XIV *Back to Communication Theory*

SURELY, IT IS WONDERFUL if a new idea contributes to the solution of a broad range of problems. But, first of all, to be worthy to notice a new idea must have some solid and clearly demonstrated value, however narrow that value may be.

An information theorist has criticized me for exploring in this book possible applications of information theory in fields of language, psychology, and art. To him, the relation between such subjects and information theory seems marginal or even dubious. Why distract the reader from the clearly demonstrated value and importance of information theory by discussing matters concerning which no clear value or importance can be demonstrated?

Partly, in writing this book I have felt an obligation to the reader to discuss relations between information theory in its solid and narrow sense and various fields with which it has been connected in the writings of others. Partly, I believe, that information theory is useful in helping us in talking sense or at least in keeping from talking nonsense in connection with some linguistic, artistic, and psychological problems. However, there is a danger in overemphasizing such matters in a book on information theory.

It would certainly be wrong to assert or to believe that information theory is valuable chiefly because of wide-ranging connections

CHAPTER I *The World and Theories*

In 1948, Claude E. Shannon published a paper called "A Mathematical Theory of Communication"; it appeared in book form in 1949. Before that time, a few isolated workers had from time to time taken steps toward a general theory of communication. Now, twelve years later, communication theory, or information theory as it is sometimes called, is an accepted field of research. Several books on communication theory have been published, and several international symposia and conferences have been held. The Institute of Radio Engineers has a professional group on information theory, whose learned *Transactions* appears quarterly, and the journal *Information and Control* is largely devoted to communication theory.

All of us use the words communication and information, and we are unlikely to underestimate their importance. A modern philosopher, A. J. Ayer, has commented on the wide meaning and importance of communication in our lives. We communicate, he observes, not only information, but also knowledge, error, opinions, ideas, experiences, wishes, orders, emotions, feelings, moods. Heat and motion can be communicated. So can strength and weakness and disease. He cites other examples and comments on the manifold manifestations and puzzling features of communication in man's world.

Surely, communication being so various and so important, a

theory of communication, a theory of generally accepted soundness and usefulness, must be of incomparable importance to all of us. When we add to *theory* the word *mathematical,* with all its implications of rigor and magic, the attraction becomes almost irresistible. Perhaps if we learn a few formulae our problems of communication will be solved, and we shall become the masters of information rather than the slaves of misinformation.

Unhappily, this is not the course of science. Some 2,300 years ago, another philosopher, Aristotle, discussed in his *Physics* a notion as universal as that of communication, that is, motion.

Aristotle defined motion as the fulfillment, insofar as it exists potentially, of that which exists potentially. He included in the concept of motion the increase and decrease of that which can be increased or decreased, coming to and passing away, and also being built. He spoke of three categories of motion, with respect to magnitude, affection, and place. He found, indeed, as he said, as many types of motion as there are meanings of the word *is.*

Here we see motion in all its manifest complexity. The complexity is perhaps a little bewildering to us, for the associations of words differ in different languages, and we would not necessarily associate motion with all the changes of which Aristotle speaks.

How puzzling this universal matter of motion must have been to the followers of Aristotle. It remained puzzling for over two millennia, until Newton enunciated the laws which engineers still use in designing machines and astronomers in studying the motions of stars, planets, and satellites. While later physicists have found that Newton's laws are only the special forms which more general laws assume when velocities are small compared with that of light and when the scale of the phenomena is large compared with the atom, they are a living part of our physics rather than a historical monument. Surely, when motion is so important a part of our world, we should study Newton's laws of motion. They say:

1. A body continues at rest or in motion with a constant velocity in a straight line unless acted upon by a force.
2. The change in velocity of a body is in the direction of the force acting on it, and the magnitude of the change is proportional to the force acting on the body times the time during which the force acts, and is inversely proportional to the mass of the body.

3. Whenever a first body exerts a force on a second body, the second body exerts an equal and oppositely directed force on the first body.

To these laws Newton added the universal law of gravitation:

4. Two particles of matter attract one another with a force acting along the line connecting them, a force which is proportional to the product of the masses of the particles and inversely proportional to the square of the distance separating them.

Newton's laws brought about a scientific and a philosophical revolution. Using them, Laplace reduced the solar system to an explicable machine. They have formed the basis of aviation and rocketry, as well as of astronomy. Yet, they do little to answer many of the questions about motion which Aristotle considered. Newton's laws solved the problem of motion as Newton defined it, not of motion in all the senses in which the word could be used in the Greek of the fourth century before our Lord or in the English of the twentieth century after.

Our speech is adapted to our daily needs or, perhaps, to the needs of our ancestors. We cannot have a separate word for every distinct object and for every distinct event; if we did we should be forever coining words, and communication would be impossible. In order to have language at all, many things or many events must be referred to by one word. It is natural to say that both men and horses run (though we may prefer to say that horses gallop) and convenient to say that a motor runs and to speak of a run in a stocking or a run on a bank.

The unity among these concepts lies far more in our human language than in any physical similarity with which we can expect science to deal easily and exactly. It would be foolish to seek some elegant, simple, and useful scientific theory of running which would embrace runs of salmon and runs in hose. It would be equally foolish to try to embrace in one theory all the motions discussed by Aristotle or all the sorts of communication and information which later philosophers have discovered.

In our everyday language, we use words in a way which is convenient in our everyday business. Except in the study of language itself, science does not seek understanding by studying words and their relations. Rather, science looks for things in nature, including

our human nature and activities, which can be grouped together and understood. Such understanding is an ability to see what complicated or diverse events really do have in common (the planets in the heavens and the motions of a whirling skater on ice, for instance) and to describe the behavior accurately and simply.

The words used in such scientific descriptions are often drawn from our everyday vocabulary. Newton used force, mass, velocity, and attraction. When used in science, however, a particular meaning is given to such words, a meaning narrow and often new. We cannot discuss in Newton's terms force of circumstance, mass media, or the attraction of Brigitte Bardot. Neither should we expect that communication theory will have something sensible to say about every question we can phrase using the words communication or information.

A valid scientific theory seldom if ever offers the solution to the pressing problems which we repeatedly state. It seldom supplies a sensible answer to our multitudinous questions. Rather than rationalizing our ideas, it discards them entirely, or, rather, it leaves them as they were. It tells us in a fresh and new way what aspects of our experience can profitably be related and simply understood. In this book, it will be our endeavor to seek out the ideas concerning communication which can be so related and understood.

When the portions of our experience which can be related have been singled out, and when they have been related and understood, we have a *theory* concerning these matters. Newton's laws of motion form an important part of *theoretical physics,* a field called *mechanics.* The laws themselves are not the whole of the theory; they are merely the basis of it, as the axioms or postulates of geometry are the basis of geometry. The theory embraces both the assumptions themselves and the mathematical working out of the logical consequences which must necessarily follow from the assumptions. Of course, these consequences must be in accord with the complex phenomena of the world about us if the theory is to be a valid theory, and an invalid theory is useless.

The ideas and assumptions of a theory determine the *generality* of the theory, that is, to how wide a range of phenomena the theory applies. Thus, Newton's laws of motion and of gravitation

are very general; they explain the motion of the planets, the timekeeping properties of a pendulum, and the behavior of all sorts of machines and mechanisms. They do not, however, explain radio waves.

Maxwell's equations[1] explain all (non-quantum) electrical phenomena; they are very general. A branch of electrical theory called *network theory* deals with the electrical properties of electrical circuits, or networks, made by interconnecting three sorts of idealized electrical structures: resistors (devices such as coils of thin, poorly conducting wire or films of metal or carbon, which impede the flow of current), inductors (coils of copper wire, sometimes wound on magnetic cores), and capacitors (thin sheets of metal separated by an insulator or dielectric such as mica or plastic; the Leyden jar was an early form of capacitor). Because network theory deals only with the electrical behavior of certain specialized and idealized physical structures, while Maxwell's equations describe the electrical behavior of any physical structure, a physicist would say that network theory is *less* general than are Maxwell's equations, for Maxwell's equations cover the behavior not only of idealized electrical networks but of all physical structures and include the behavior of radio waves, which lies outside of the scope of network theory.

Certainly, the most general theory, which explains the greatest range of phenomena, is the most powerful and the best; it can always be specialized to deal with simple cases. That is why physicists have sought a unified field theory to embrace mechanical laws and gravitation and all electrical phenomena. It might, indeed, seem that all theories could be ranked in order of generality, and, if this is possible, we should certainly like to know the place of communication theory in such a hierarchy.

Unfortunately, life isn't as simple as this. In one sense, network theory is less general than Maxwell's equations. In another sense,

[1] In 1873, in his treatise *Electrictity and Magnetism,* James Clerk Maxwell presented and fully explained for the first time the natural laws relating electric and magnetic fields and electric currents. He showed that there should be *electromagnetic waves* (radio waves) which travel with the speed of light. Hertz later demonstrated these experimentally, and we now know that light is electromagnetic waves. Maxwell's equations are the mathematical statement of Maxwell's theory of electricity and magnetism. They are the foundation of all electric art.